MIR: NP-Structures

(draft version)

Status: Proposal
Date: 2.1.90
Author: G.Thurmair

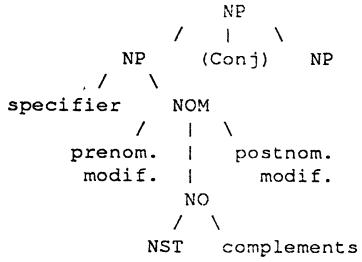
(This paper should be read in combination with Meya's paper on NP structures. Some parts are explained in more detail there).

0. General remarks

An NP consists of a head, its complements, modifiers, and specifiers. The head is either a lexical element, or a complex structure. Complements are arguments of the head; they either are taken over from the verb, or are lexically specified. Modifiers can be prenominal or postnominal and will be marked as such. Specifiers can be determiners or quantifiers. The NP as a whole can be modified.

In MT, in addition, we have to cope with terms (i.e. terminological units, consisting of several elements like nouns + PPs, ADJs + Nouns, etc.).

According to the literature, the basic layout of an NP would be:



As the terminal units of the NP can be (morphologically) complex, we would end up in more than three bars for an NP.

There are several changes in this basic structure wrt METAL. They will be described in the following sections.

1. Types of heads

The head of an NP can have special properties and need special attention. The simplest case is that a head is a simple lexical unit, like "house", "tree" etc.; but there are other cases as well.

A head can be complex: it can be result of a process of derivation or compounding. This has to be considered by a MT system. Moreover, the head can consist of a constant (which is either lexical or not). Constants need special attention, too.

The following chapters deal with these special cases of heads.

1.1 Compounds

1.1.1 General remarks

The head elements of an NP, the nouns, can be lexical (i.e. simple, like "house", "tree") or complex. Complex nouns are derivations or compounds. This chapter deals with compounds.

In principle, compounding is a means to create new lexical units from already existing lexical material. Therefore, they function as basic units in their syntactic environment just like non-compounds. This means that in principle, they cannot be analysed compositionally, e.g.

(1) GE: Kind-s-kopf (silly person)
Kind-er-kopf (head of a child)
(2) GE: Wasser-s-not (too much water)
Wasser-not (not enough water)
(3) GE: Jaeger-schnitzel
Schweine-schnitzel
(4) GE: Wald-meister (...)
(5) SP: llave maestra
(6) SP: boca calle
(6) EN: house wife

Compounds undergo a process of lexicalisation which means that some of them have to be coded entirely in the lexicon.

On the other hand, as compounding is one of the basic tools for the ad-hoc creation of new units, these units have to be understood, i.e. there must be means "decomposing" ad-hoc compounds. This gives us the chance to do compound analysis. Also, compounds influence the behaviour of the noun in tis syntactic environment (e.g. argument selection, determination, etc., see below).

On the other hand, other languages have other means for creating new lexical units: They use adnominals, PP-attachment etc. Now, complex semantic units can be compounds in German (7), nominal attributions in English (8), nouns with prepositionals in Spanish or French (9).

(7) DE Schweine_schnitzel
(8) EN porc escalope
(9) SP escalopa de cerdo
FR escalope du porc

From these considerations, it follows that it is very difficult to define what a "compound" really is. It seems that it is a kind of semantic unit consisting of a (syntactically or morphologically) describable structure. But the same holds for the description of multiwords in other languages. This seems to indicate that the notion of "compound" is morphologically determined and therefore not too relevant; and what is relevant is the complex semantic unit which we have to consider. (And in fact, German compounds incorporate complex syntactic properties, see below.)

From this, it follows that the distinction between compounds and multiwords is not possible in MIR; and we need a common representation of the structures in question whether something is a compound or a multiword. Compound and multiword is just the morphological counterpart of a common semantic property.

What is much more important, however, is whether such a "complex semantic unit" is lexicalised or compositional. Criteria for lexical treatment of compounds are given in Neunzig/ Grauwinkel 1988.

There is a consequence of this in METAL: There is no simple borderline between morphology and syntax: A (morphological) compound has to be analysed as a tree which has to be transformed into another (maybe syntactic) tree in transfer. Technically spoken, the rule type COMPOUND has to be treated differently: Whereas in GE thea are word rules (and used for morphological analysis e.g. in preanalyse), in EN they are syntax rules. Therefore, we propose to eliminate "compound" as a possible value of the rule type, and have just the distinctioon between "morphological" and "syntactic" as possible rule types.

The problem then is to have an interface between morphology and syntax: Morphology cannot output just unstructured terminal elements (as compounds must have an internal structure); (and therefore a rule type "COMPOUND" might be useful).

1.1.2 Analysis of compounds

Inside of compounds, we should have a kind of structure: We should attach the FLEX element to the compound as a whole, i.e. as last step in the derivation. This can be seen from examples like (10) where the flex belongs to the coompound as a whole rather than its rightmost element (as "up" is not inflected).

(10) EN: pick-up-s

However, as we treat compounds and multiwords alike, we propose to attach the FLEX to the rightmost element before we go on further; i.e. (11) and (12) will have similar structures (11') and (12'):

(11) GE: Haus-Tuer-en
(11')

```
NST   NST NFLEX
 |    --------
NO      NO
----------
   NO
```

(12) EN: disk drive-s
(12')

```
NST  NST  NFLEX
 |   ----------
NO       NO
-----------
     NO
```

(This is not so important, however, as the FLEX node disappears anyway in MIR, and the FLEX info can be copied down to the first right NO before transfer. However, it makes transfer easier, e.g. into Spanish, cf. (13):

(13) DE: Frauenklinik -> SP: clinica para senoras

Below this level, we need to know which element is the head of the compound, and which is the specifier. And we need to know some more details of the construction we are building.

1.1.2.1 Head and specifier

Most compounds will have a head and specifier. They will be marked as such, using the FUNC feature (FUNC like ROL describes the function of these compound parts, but we cannot use ROL here, see below):

FUNC HEAD marks the head
FUNC SPEC marks the specifier

The syntactic category of the SPEC should not be NST/AST/VST but rather NO/A/V, as the following examples show:

(14) GE: naechst-gelegen (A)
(15) GE: Goetter-Speise (N)
(16) GE: Ab_bau-genehmigung (V)
(17) EN: humanities reasearch (N)

I.e. the specifiers can be inflected etc. as well.

In multi-part-compounds, elements can be heads and specifiers at the same time; and there can be very many internal structures, cf.

(18) DE: elektro rasen maeh maschine

```
                  spec   head
                  -----------
            spec      head
           --------------
spec          head
-----------------
```

(19) DE: fertig teil fabrikats erzeugung

```
spec  head
----------
    spec      head
    --------------
          spec          head
          ------------------
```

(20) DE: sonder druck buch handel

```
spec   head spec  head
----------- ----------
     spec       head
     ---------------
```

As we want to treat compounds like multiwords, we will try to have a flat structure also for multi-part compounds. Therefore we must mark the exact status of each constituent. This is done by an additional

feature which marks the CAN of the respective head of a constituent:
(18) would have (18'), (19) (19'), (20) (20'), etc.

```
(18')  elektro        rasen          maeh        maschine
         NO             NO            VB            NO
       FUNC SPEC     FUNC SPEC      FUNC SPEC     FUNC HEAD
       HEAD maschine HEAD maschine  HEAD maschine    -

(19')  fertig      teil           fabrikats       erzeugung
         A           NO             NO              NO
       FUNC SPEC   FUNC SPEC      FUNC SPEC       FUNC HEAD
       HEAD teil   HEAD fabrikat  HEAD erzeugung      -

(20')  sonder     druck        buch         handel
       FUNC SPEC  FUNC SPEC    FUNC SPEC    FUNC HEAD
       HEAD druck HEAD handel  HEAD handel   -
```

With these two features we seem to be able to model the internal structure of a compound, relying on the precedence order of its parts.

In GE (and EN and some other languages), compounds underly the righthand head convention, i.e. head of the compound is its rightmost element. Nevertheless, this does not hold for all languages (e.g. GR has right as well as left headedness). Therefore, we must find out in these cases where the head is, and mark it accordingly. Non-terminal nodes will be marked with the feature sets of their heads, with the additional FUNC marker.

As (18) - (20) show, we could have many ambiguities in compounds with more than two parts. If we cannot resolve it (by looking at the CATs used, looking at deverbality etc.), it is proposed to group three-part compounds like (21)

```
(21)      morph1  morph2  morph3
          spec     head
          ------------
                spec      head
                ----------
```

This seems to cover more cases than strict right-to-left-order, at least in GE (see the evaluation of Neunzig/Grauwinkel 88 for GE). Everything else may be forced into right-to-left order. It should be noted that this ordering does influence the translations, see again Neunzig/Grauwinkel with examples:

(22) [[Zylinder kopf] schraube]
(23) [[Verteiler schnecken] zeichnung]

1.1.2.2 Categories of compound specifiers

Compounds always end up to be NOs.
We assume that the heads be NSTs. These NSTs can be result of lexicon lookup, or of derivations. If they are produced by derivational processes then there might be restrictions to the process of compounding (e.g. compounding of nominalised verbs etc.). The heads are combined with FLEXes to form NOs. This is the level where compound attachment is done.

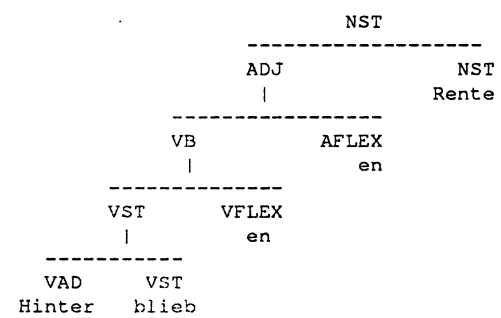
Specifiers can have several categories, however: They can be NSTs, ASTs, VBs etc. As this choice determines the translation sometimes (cf. 24), we cannot just take one reading in the way of morphological segmentation (as proposed by Khan/Oppenheim 88) but keep several segmentations and leave it to analysis to create the proper interpertation:

(24) GE: Druck (NO) - Programm -> EN: print program
     GE: Druck (V) - Programm -> EN: printing program

The categories of the specifiers will be nonterminal again: A verbal specifier can be prefixed (cf. (25)), a AST can be formed from a participle (cf. (26)) etc.:

(25) GE: Ab_bau-Genehmig_ung (VAD - VST - NST - NSUF)
(26) GE: Hinter_blieb_en-en-Rente (VAD - VST - VFLEX - AFLEX - NST)

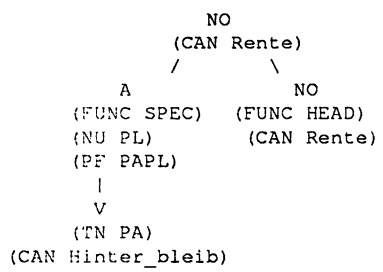
Therefore, the specifier categories should be nonterminal as ell, i.e. A, NO, V, instead of AST/NST/VST. Moreover, we should not enumerate all the possible sequences of these specifier categories (as done in the GE grammar) but really build nonterminals. We will have 0-bar types of categories (i.e. results of derivation and compounding, also inflection!, but no complements or modifiers etc.: A, NO, V). (26) in GE analysis would be represented as (26'):



This takes into account that there might be inflection (also comparation) inside of a compound, e.g. (27), (28):

(27) GE: meist-bietend
(28) EN: savings account

Here, as elsewhere in MIR, the FLEX nodes will be removed and their respective information is copied to the next higher nodes. MIR structure of (26') therefore will be (26''):



As a result, compounds will be determined by an NO, and will have a flat structure below them, formed from A,NO, and V nodes. One of them is marked to be the head, the others are Specifiers, the head of which is stated explicitly.

1.1.2.3 Infixes

Some languages, like GE, have special infixes for compound formation. They are idiosyncratic (i.e. have to be coded in a special LINK feature of a lexical entry). They will be removed at MIR level as no specific function can be attached to them. (if they can be used for disambiguation, this will influence the respective "main" categories).

1.1.2.4 Roles and Arguments

Compound elements can be relevant for the argument structure of the whole NP. This is the case if the head is verbal or deverbal. Then an argument of this verb might be attached in a compound position. In (29), the FUNC SPEC is the $DOBJ of the head, in (30), it is the $SUBJ:

(29) GE: Staedte-Zerstoerung
(30) GE: Papst-Reise

To cover this, we must tell the outside NP that a role has already been realised. To do this, we need the feature ROL (which therefore cannot be used for FUNC).

This is a problem for NP framing: Framing presupposes that the complements are accessible in a flat order below the node to be framed.

This would mean to dissolve the compound structure and flatten its parts below the NP node. But this contradicts to the fact that compounds behave as units in the NP structure.

Therefore we propose to follow a different strategy: Determine the role locally inside the compound (e.g. using features like deverbal, TT etc.), and mark the head as having a specific role. NP framing will be called with "NO-$SUBJ" etc. (depending of what has already been realised). (Usually, however, this role cannot be determined uniquely; there are only few really clear cases. But IF we can determine it, we can express this behaviour).

1.1.2.5 Semantics

Specifiers can help to disambiguate TYs of the heads (and vice versa). This should be treated in connection with the overall review of semantics. However, we need the TYs of the other compound elements, as pointed out in Neunzig/Grauwinkel 88. Therefore, they are copied as STY and HTY respectively. This is due to the fact that transfer is done on NST (terminal) level in the current system. If we have transfer on NP level as well (as proposed below), this copying needs not be done.

1.1.2.6 Coordinative compounds

There are elements with two heads, (coordinative compounds), e.g. (31) or (32):

(31) GE: nass-kalt
(32) GR: politiko-oikonomikos
(33) GE: Hosenrock

1.1.2.7 Gapping of compound elements

In some languages (like GE), we can have gapping od the compound heads, cf. (34):

(34) GE: Haus- und Gartentuer
(35) GE: gruene Haus- und rote Gartentuer
(36) GE: die rote Haus- und die gruene Gartentuer
(37) GE: Blumenbeet und -hecke

(34')

Haus	[tuer]	und	Garten	tuer
NST	NST		NST	NST
NO	NO		NO	NO
	(GAPPED T)			
	NO		NO	
	NO			

In (34), we should have a conjunction of NOs, producing a structure (34') where the first "Tuer" will be marked as "GAPPED T". In (35), we will also have a conjunction of NOs, and in (36), there will be a conjunction of NPs. In neither case, conjunction will be handled within the compounds. This also holds in cases like (37) where gapping of the specifier occurs: This is a conjunction of NOs.

This proposal makes transfer pretty easy:

(38) DE: Haus- und Gartentuer -> SP: puerta de la casa y del jardin

The specifying NOs are consistently transformed into PPs (see also paper of Meya on this topic).

1.1.3 Transfer of compounds

In compound transfer, there are the following alternatives:

If the compound is coded as a whole, transfer is done for this lexical unit. To keep the lexicon consistent, all other languages have to give the transfer as well. This might lead to redundancies in the dictionaries.

Therefore, we should give the compound transfer (if the compound is

non- lexicalised) only in the respective transfer-lexica. The means we have are the features COMPLOC (now called FUNC) (specifying different translations for lexical units depending on whether they are heads or specifiers in a compound), STY/HTY (i.e. the semantic type of head or specifier), and we should add the possibility of specifying specific SCAN/HCANs (i.e. "if head is XY and specifier CAN is YZ then translate ABXC"). Examples are

```
(39)        Eisen   ->      planta siderurgica
            FUNC SPEC
            (REQ @HEAD CAN Huette)

(40)        Feuer   ->      mechero
            FUNC SPEC
            (REQ @HEAD CAN Zeug)
```

This reduces coding effort to only transfer lexica and makes compounding a matter of two languages rather than all languages. But it leads to redundancies if in a specific combination, both head and specifier translate in a specific way; cf. in (41) the test has to be given twice, both for head and specifier:

```
(41) ABC        ->     CDE
     Tests:  FUNC HEAD
             REQ @SPEC CAN BCD

     BCD        ->    TZU
     Tests:   FUNC SPEC
              REQ @HEAD CAN ABC
```

In addition, we need the possibility of adding lexical material, e.g. specific prepositions, adjectives etc.; examples are (42), (43), cf. Neunzig/Grauwinkel 88

```
(42) Schnecke ->    tornillo
    Tests: FUNC HEAD
           REQ @SPEC CAN Verteiler
    XMFS: ADD (ALO sin_fin_del_delco)

(43) Huette ->      planta
    Tests: FUNC HEAD
           REQ @SPEC CAN Eisen
    XFMS: ADD (CAN siderurgico) (CAT AST) (PO AFT) ...)
```

We propose to allow for transformations in the NST-transfer here as well, e.g. have something like

```
XFR-ADD (:VALUES TL-CAN TL-ROL TL-CAT ... )
```

(This would be a more general proposal than the ones proposed by Neunzig/Grauwinkel (who have ADD-STRING-CAn, ADD-STRING-TY, ADD-STRING-PREP, etc.) and Gebruers (who proposes a REST feature containing an ALO), cf. Gebruers 89).

This transformation should be called at NP level, however, not at NST level, just as the FRX does for verbs. The principle should be that we do transfer of controllers of each category first, then transfer of the controllees. This implies the application of the TRF mechanism to other categories.

1.2. Derivations

NPs can have heads which are results of a process of morphological derivation. In the METAL languages, derivation means affixation but also category changes (we leave out phenomena like Umlaut/Ablaut).

In principle, for derivation, the same strategy holds as for compounding: As far as a derivational process can be described as rule-based, and as far a rule-based transfer can be given, METAL should try to capture these generalisations. All other cases have to be put into the lexicon.

There is one caveat, however: Derivational analyses tend to blow up the process of morphological segmentation, and it is not easy at all to restrict it to the level we want (see Thurmair 87 for a more detailed discussion). The reason is that we have very short morphemes which are recognised nearly everywhere, and that the constraints to be applied

are difficult to state especially if there is combined prefixing and suffixing.

1.2.1 Affixation

Tests with affix analysis mainly have been done in the EN analysis grammar. There is some uncertainty how this should been done, wrt to analysis as well as transfer.

Considering suffixtion alone, the suffixes form the head of such a construct from a morphological point of view (as they determine the part of speech, the semantic type, sometimes the argument type, etc.). However, what is the rest: specifier? argument? How can it be treated?

Transfer is even more problematic: If something is regular in one language (like (44)-(45)), it is not clear at all if a parallel regularity exists in a target language. But this is a prerequisit for a proper translation.

(44) EN: resist-able
(45) IT: (cort)-INO vs (cort)-ONE

For these reasons, it is recommended to use derivational analyses only as fallbacks for incomplete dictionary entries.

1.2.2 Category change

Here we mean nominalisation of verbs or adjectives. This can be achieved by adding the respective flex, by adding a determiner, etc.; cf. (46)-(48); (note that (49) is a real case of a CLS-SUB).

(46) GE: das Verstehen (NST)
(47) GE: laufend-e (AST)
(48) EN: the running of the program
(49) EN: his running the program

These processes have influences on the argument changes (esp. nominalisation of verbs, see below). Nominalised constituents are marked as such, using the boolean features (NVRB T), (NADJ T), (NADV T). They might be translated in a different way, cf. (49). Here, we might copy up the features PF as well (gerund):

(49) GE: laufen -> EN: run
GE: Das Laufen -> EN: running

It has to be taken care that the features needed for a well-formed noun are constructed from the other constituent; some may undergo changes, some have to be newly constructed (e.g. TYN).

(Gerunds in EN are not treated as such a kind of nominalisations; they enter the NP on a higher level, after complementation and modification, see below).

1.3 Constants

Constants differ from other lexical units as follows:

The way they enter the grammar is different: Some are in the lexicon (namely when they have to be translated into something different, cf. (50), (51):

(50) GE: UdSSR -> EN: USSR
(51) GE: EDV -> EN: EDP

In this case, they are in the lexicon, marked as CONST T.

Some others are put into METAL constant braces. They are not lexically analysed up to now, but might have to be (depending on the new text processing software). Everything between such braces is a constant by definition.

Finally, they enter the grammar as UNKnowns with special ORthography information and are turned into CONSTs by rules.

However they enter the system, all constants are uniformly marked as (CONST T). (What a constant is is described in Thurmair 89). Grammar turns them into nouns where appropriate, because syntactically, they behave like nouns: they can be modified, specified, etc.

Within the morphology however, constants differ from other "regular" nouns in their morphological behaviour: They don't simply take flexes, they do not take affixes usually, they cannot frely combined with other lexical units in compounding. This behaviour causes severe problems in morphological segmentation if CONSTs are treated as NST in morphology, cf. Thurmair 89.

CONSTs can be marked as ABBs if people want to do so. (ABBs are just important in case the control ALO selection in transfer).

1.4 Abbreviations, Units and others

Abbreviations are nouns. Their treatment is described in Wiesner/ Jezierny 89. They are marked as ABB T. As they sometimes allow for final PNCT and sometimes they do not, this behaviour has to be coded as (FPNCT OBL/OPT) (see minutes of Barcelona workshop 89).

Abbreviations are treated as ALOs of their long forms. While the FPNCT feature is not relevant in MIR (just in analysis, where PNCT attachment should be treated like FLEX attachment, and in generation, where a PNCT is added or not), the ABB feature is relevant as it might control ALO selection.

The same holds for unit descriptions like "m", "kg", "$" "%" etc. They have to be marked as (TYN MEA) in the lexicon as they are important in the construction of measure NPs. As some of them (e.g. "km" as opposed to "%") also specify the dimension of measure, they could be marked as both TYN MEA and TYN LOC. This will not cause problems if the proposal of readings for TYNs is adopted, where sets of features can be specified. Syntactically, they behave like usual nouns and can be heads of NPs, fillers of MEA-SADVs etc., cf. (22)-(24):

(52) EN: He bought up to 5 kg
(53) DE: Er wiegt 211 kg
(54) DE: Der Preis steigt um 5 %

Unit descriptions should be marked as (CAT NST) (CAN percent) (ALO %) (ABB T) (TYN MEA) as they are abbreviations of a unit description. In doing so we can use the ABB mechanism described in Wiesner/Jezierny 89.

1.5 Unknowns

Unknowns must go into NSTs. In this process, we attach all the obligatory features together with all their values as long as we do not have online defaulting (see Adriaens 90 on this topic). We should mark unknowns with UNK T.

2. Complements and roles

Nsts leave the morphological level and become NOs by adding flexes (or the 0-FLEX). Flexes copy their respective feature set onto the NO node, update some of the NST features, and are then deleted on the way to MIR.

The first job of a newly built NO is to look out for possible complements. This is different from multiword terms as proposed in Vazquez 88; in (1), the ADJ would not be a role but part of a term; whereas in (2), the PP is a role (but not necessarily part of a term).

(1) DE: blinder Passagier -> SP: polizon
(2) DE: Durst auf Bier -> SP: sed de cerveza

Terms are treated below. Problems with roles are described here.

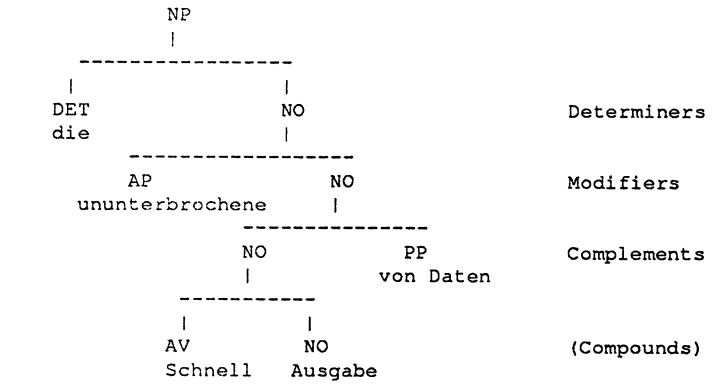
2.1 Noun Arguments

Noun arguments are the first constituents to be attached to the heads

of NPs. This can be seen from the fact that they are included in the scope of possible modifiers / quantifiers / determiners:

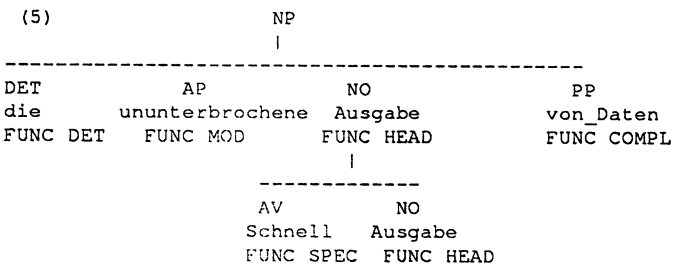
(3) EN: [the [silly [hope for peace]]]
(4) DE: [die [ununterbrochene [Schnell-Ausgabe von Daten]]]

This contradicts to the standard METAL analysis where PPs are not attached before NP level. Here, they would be attached at NO level already:



This analysis reflects the scope relations in the tree structure: Complements are also complements of compounds, modifiers modify heads plus complements, etc.

However, in MIR, we will end up with flat structures like (5). Therefore we must mark the respective functions of the NP constituents. We will do this using the FUNC feature:



The FUNC feature will have the values SPEC HEAD COMPL MOD APP (for apposition, see below) DET.

Now, within complements we can have different roles. This is what the ROL feature is about. It will have (a subset of) the "classical" METAL values: $SUBJ $DOBJ $POBJ $SOBJ $LOC $TMP $MEA. We have to distinguish between regular and deverbal nouns.

2.1.1 Regular nouns

We must distinguish between deverbal and non-deverbal nouns. The former take their ARGS into the new context, the latter might have ARGS as well. In both cases, the ARGS differ from the verb ARGS in that they are optional. It follows that it is very difficult to give tests for coding ARGS for "regular" (i.e. non-verbal) nouns. Cases in (6) are generally considered to be ARGs, cases in (7) are far less clear.

(6) EN: hope FOR sth
GE: Mitteilung AN
Druck AUF
Drohung MIT
SP: preposicion de regimen
...

(7) EN: output TO printer
salida a / de
...

The current METAL system treats these cases with form-of-complement and marker-of-complement features. This is not sufficient as is allows only for one complement (with just one preposition in case of PP).

The proposal is to mark constituents as complements in the following cases:

1. If there is an attributive clause which might depend on the noun

(8) GE: Hoffnung, es zu tun
Angst, zu spaet zu kommen

This should be done as otherwise, CLS-SUBs are tried to be attached to the clause as a whole. (This is more a practical reason: Do clause-attachment of CLS-SUBs except they are marked on nun level). These nouns have ROL value $SOBJ.

2. If there is a PP which requires a specific preposition, like the ones above. There can also be a set of them, like in

(9) GE: Urteil gegen / ueber / zu
Hilfe an / fuer

They should be coded with ROL $POBJ and a certain prep (and case). This is a grey area, however, and users should code POBJs if they feel that a PP occurs relatively often with this noun; and if it occurs, it should be attached to the noun rather than the matrix clause.

3. If there can be a genitive NP, we would like to know whether it is the subject or the object of the noun (genitivus subjectivus / objectivus); cf.

(10) GE: Uebersetzung des Textes ($DOBJ)
Festnahme des Bankiers ($DOBJ)
Ruf des Vaters ($SUBJ)
Niederschrift des Kollegen ($SUBJ)

(Is this relevant in MT? Can it be determined monolingually?)

As a whole, we will have the following ROL values for ARGS of non-verbal nouns: $SUBJ, $DOBJ, $POBJ, $SOBJ, $TMP, $LOC $MEA; these values can have additional subcategories.

As these roles can be realised differently (e.g. as specifiers of a compound) we do not want to merge them with other information. Therefore we have a feature describing the FUNCTION of a constituent, in terms of HEAD, SPEC, etc., and the ROL feature describing just the status of a possible complement. Roles only desribe possible complements: They mark a subset of FUNCs, namely the FUNC COMPL nodes (but roles may occur outside of FUNC COMPL nodes, see below).

2.1.2 Deverbal nouns

They inherit their verbal ARGS, but in a very specific way. (If we want to express this fact, we need the same values for the ROL feature at NP level).

(11) GE: Das Programm verarbeitet Daten
Die Verarbeitung von Daten durch das Programm

(12) GE: Der Bankier bezieht Rente
Der Bezug von Rente durch den Bankier
Das Beziehen von Rente durch den Bankier
Des Bankiers Beziehen von Rente

The standard transformation seems to be that the $SUBJ of the clause goes into a genitive NP or a passive PP, $DOBJ goes into the "von"-PP. The respective constituents therefore can be marked with the respective roles. The roles which may occur here are $SUBJ, $DOBJ, $POBJ, and maybe others.

(see paper of Meya for other examples).

2.2 Location of roles in a NP

If we want to apply standard framing procedures to NPs, we face the following difficulty: In verb framing, all potential arguments may occur one level below the CLS level, on a flattened tree. Here in NPs, complements / arguments can be found on very different levels:

They can be inside of the head nouns, in the case of compounds, cf.:

(13) GE: Drohung MIT Mord -> Morddrohung ($POBJ)
Reichtum AN Kohle -> Kohlereichtum ($POBJ)
Reise DES Papstes -> Papstreise ($SUBJ)

(This is relevant for preposition insertion e.g. in the case of generation into Spanish). Here we have FUNC SPEC.

They can be attached at NO-level, as PPs, CLS-SUBs etc.; this is the standard case; here we have FUNC COMPL.

They also can be attached at specifier level, if they are represented as genitives:

(14) EN: John's destruction of the city

Here, we have FUNC DET.

Nevertheless, in all those cases roles can be expressed; therefore we need two features for ROL and FUNC, as explained above. All these relations can be expressed by using the FUNC and ROL feature appropriately.

2.3 NP framing

NP framing differs from PRED framing in two respects: First, it cannot be used as a test, i.e. it will never be the case that a NP reading is rejected due to framing. This is due to the second respect: All possible roles in a NP are optional.

For framing, this means that the only task is to identify possible ARGS if they are realised. To do this, it is not necessary to compare different frames, to check and reject combinations, etc.; framing in NPs restricts to identifying possible roles if they are realised. The only concern is that roles should not be attached twice within an NP.

But this means that framing is just a local operation which checks whether a given constellation is relevant for being marked with a ROL feature. In this respect, we do not need a global framing procedure; we just need

- a comparison whether a constituent found matches a certain ARGS pattern in the entry of the head noun
- a marking mechanism for the ARG found (ROL plus the respective value)
- a check to make sure that roles are not used twice.

Result of the framing of NPs will be a list of assigned roles (stored in a feature called ASF). This list can be updated by compounds as well as real complements or specifier genitives.

2.4 Transfer of NP complements

As there is just optionality of all NP complements, transfer does not select between different frames, like verb transfer does. It still transfers the head with certain tests; but we should be able to express changes on the basis of the presence of certain roles (this might be needed for translation of terms anyway, see below). Moreover, we need a description of systematic changes, e.g. specific transfer of prepositions, like PP-to-PP. This is not possible in the existing system as SL MC has nothing to do with TL MC which is problematic if more than one complement is marked in the lexicon. Also, we need some more complex transformations, e.g. for cases like

(15) GE: die Moeglichkeit, das zu tun -> EN: possibility of DOING it

This change must be triggered by the ARGS of the head; therefore we need a possibility to do transfer on NP level (a kind of FRX) which is responsible for the whole NP. Then we have to pass down the transferred

head using the TRF feature, and at the same time, we have to mark possible complements (and modifiers) for structural changes if necessary. (This possibility is needed for term translation also, see below). The result would be an entry like (16):

(16) DE: Moeglichkeit -> posssibility
Tests: (REQ @$SOBJ ICP INF+)
XFMS: (MAP $SOBJ ICP INF+ -> $SOBJ ICP ING)

This would be done at NP level (requiring a flat NP structure, however). MAP would mark the SOBJ with the ICP value required, and the head would be marked with the TRF feature. (In fact, the test in (16) does not make sense as the role is optional anyway, and "Moeglichkeit" should go to "possibility" regardless if a $SOBJ is realised or not).

(17) DE: Hoffnung -> hope
Tests:
XFMS: (MAP $POBJ Prep auf -> $POBJ Prep for)

(As all this role assignment is optional, we have to take care of the fact that one language marks "Hoffnung" with a $POBJ whereas another language does not. Therefore a more general MAP without roles should also be possible, as in (14):

(17) DE: Hoffnung -> hope
Tests:
XFMS: (MAP NIL Prep auf -> NIL Prep for)

To be able to do such operations, we need a flat NP structure where all complements are resembled, and as a cnsequence, all modifiers are on the same level as well.

To summarise this topic:
1. We will mark lexical entries for ARGS which describe what kind of ARG is possible here. This will replace the old MC/FC mechanism. 2. We will have features ROL at the respective nodes and procedures to identify them at the nodes in question. 3. We will not have a general frame procedure but rather a set of local operations. 4. We will do transfer on NP level, allowing nouns to have tests and XFMS ranging over the whole NP structure; XFMS may influence the structure of complements (and modifiers as well if needed).

3. Modifiers

Modifiers are built from nouns plus complements. They can consist of PPs, APs, AVPs etc.; they can be prenominal and postnominal. Modifiers are marked as FUNC MOD.

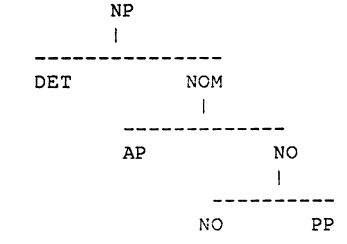
(We cannot follow the simple head-specifier distinction because we have to know what kind of specifier we have: As we have a flat NP structure there will not be a structural difference between a postnominal modifying PP and a postnominal complement PP).

3.1 Modifier structure

Modifiers are attached one level higher than complements as they modify these structures as a whole; i.e. in analysis, we have the structure (1):

(1) EN: the [crazy [hope for peace]]

(1')



In MIR, as we want to build flat structures, complements and modifiers

will be attached at the same level, below the NP; i.e. (1''):

```
(1'')              NP
                   |
        --------------------------
        DET   AP        NO        PP
            FUNC MOD  FUNC HEAD  FUNC COMPL
```

In order to be able to distinguish the functions of these constituent, we should mark the head as (FUNC HEAD) (as there can be other NPs and even NOs, see below), and the modifiers as (FUNC MOD). Complements are FUNC COMPL, and in addition have their roles. The result is that the FUNC MOD need not be unique, as other FUNCs are.

Modifiers are usually phrasal categories (XPs), with one exception (see below). But they are not grouped into just one XP of each category (as proposed in Liu 89): (2) would have structure (2'), not (2''):

(2) EN: the red fast car

```
(2')               NP
                   |
        ----------------------
        DET   AP      AP     NO
              red     fast   car
```

```
(2'')              NP
                   |
          ------------------
          DET      AP       NO
                   |
                -------
                AP    AP
```

(Structure 2'' would be appropriate if we had a conjunction of APs).

As a result, we have a flat structure with both modifiers and complements, the function of which can be determined using the FUNC feature.

3.2 Special problems with modifiers

3.2.1 Ordering

Languages can have modifiers both in prenominal and postnominal position, cf. (3)

(3) SP: el gran bloque de control de ficheros rojo

Usually, in METAL, the trees which are built are not sensitive to linear precedence, i.e. the order of nodes is not relevant. But we propose to deviate from that rule in the case of NP structures; there are two reasons for that:

1. Some languages differ in meaning between APs in prenominal and postnominal position, cf. (4):

(4) SP: una buena mujer - una mujer buena

To cover this difference, we need a PO feature (with values BEFore and AFTer head). But this is not the only problem. We also have to know the relative position of the modifiers themselves, cf. Vazquez 88: In (4), we have a "blind passenger" who is tall, whereas in (5), we have a passenger who is tall and blind:

(5) GE: der deutsche blinde Passagier
-> SP: el polizon aleman

(6) GE: der blinde deutsche Passagier
-> SP: el pasajero aleman ciego

Here we have two possibilities of representing these relations: We could mark the relative position using the PO feature as follows:

(7) the big old car from Austin owned by Roland

```
(7')            NP
                |
    ----------------------------------------
    DET    AP      AP      NO     PP     CLS-SUB
         (PO -2) (PO -1)        (PO 1)  (PO 2)
```

But this order can also be implicitly seen from the constituent order in the NP tree. Therefore we propose that the NP tree mirrors the surface constituent linear precedence, and that the PO feature just specifies whether a constituent is before (BEF) or after (AFT) the head (This is relevant for transfer).

This is important for the treatment of terms, as explained in Vazquez 88: Parts of terms are always neighbours (see below).

3.2.2 Narrow appositions or complex names

Narrow appositions are neither complements nor modifiers; they are something like identifiers, i.e. they name (or make definit) the head noun, cf.

(8) EN: the METAL system
the system UNIX V3
(9) GE: das System METAL
der Kanzler Kohl
(10) EN: the dog Fido
the car Big-Magic-Blue
(11) EN: They sold a METAL system

(8) and (9) show that positioning of these complex names can differ in different languages; therefore we need the PO feature here again. (10) shows that the property of being specified by a name is not lexical; nearly every (count?) noun can be named. The function of these names is to select an element (or a subset, cf. (11)) from a set.

The names are related with modifiers as follows: If there are other modifiers, they modify the whole name, e.g. (12):

(12) EN: they sold these beautiful performant [[METAL systems] of Siemens]

Therefore, they should be attached before the other modifiers.

Compared to complements, the names do not fit very well; both (13) and (14) are not very good ((13) being better than (14), however):

(13) GE: Hoffnung ABC auf Rettung
(14) GE: Hoffnung auf Rettung ABC

Anyway, we should mark names with a specific FUNC value (FUNC APP for apposition).

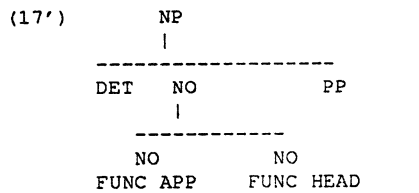
The internal structure of these identifiers can be very complex. (15) gives some examples from GE:

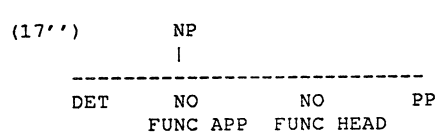
(15) GE: das System UNIX (NST)
das System UNIX 4.2 B (NO)
die Farbe Blau (AST)
die Partei Graue Panther (NO)
der Ruf "Hilfe" (NO)
der Ruf "Rette sich, wer kann" (CLS)
das Jahr 1984 (NUM)
(16) EN: They sold all the METAL systems

But apparently, we cannot have a full-fledged (i.e. determined) NP here, and on the other hand, the name falls into the scope of the overall NP determiners, cf. (16). This means that we will have a non-maximal projection to be attached to the head noun.

The question where these names should be attached can be solved in a pragmatic way: As we can attach these names on NO-level (17') or at NP level (17''), we decide for the latter alternative, as some languages can have modifiers (i.e. parts of terms) between the head and the name, cf. (18):

(17) EN: the METAL system of Siemens





(18) FR: le systeme operative UNIX

(This is another reason to specify within several NOs below the NP which one is the head).

Also, considering NUMs, the difference between (19) and (20) is (among others) captured in the FUNC feature: (19) has (FUNC MOD) whereas (20) has (FUNC APP):

(19) the year 1984
(20) 1984 years

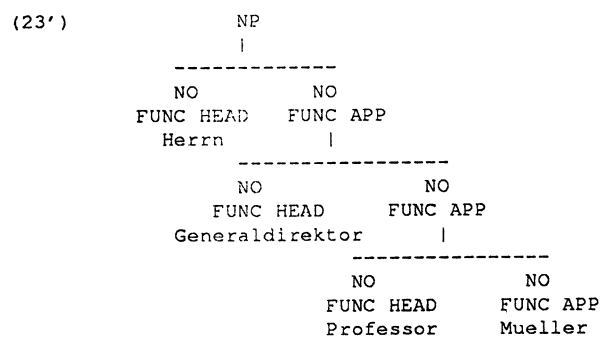
As a result, we propose to have the complex names on NP level, recognised by a FUNC MOD feature. In analysis, we should attach them before attaching complements (i.e. as very first operation of building an NP).

3.2.3 Titles

Titles are treated as complex names as well, the head of the construction being the title. At least in GE, the syntactic head is the title, cf. cases like (21), (22) where the agreement clearly includes the title, but not the name:

(21) DE: Wir treffen Herrn Mueller
(22) DE: Wir treffen den klugen Herrn Mueller
(23) DE: Wir sehen Herrn Generaldirektor Professor Doktor Mueller

Titles may be embedded into each other, cf. (23); at least in GE, the head is still the leftmost title. We propose not to flatten this structure but to treat it as one complex name, the structure therefore being (23'):



This follows form the morphological facts in DE: The respective heads are neutralised as soon as they become part of the complex name; only the remaining head is morphologically active. Flattenning would destroy these dependencies.

3.2.4 Conjunctions and gapping

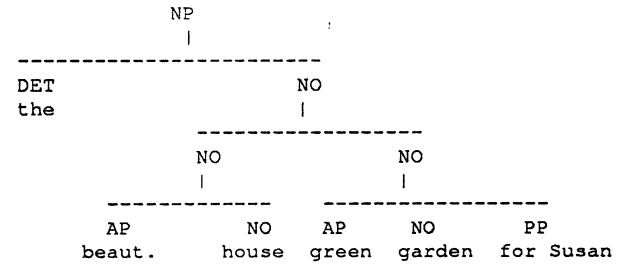
On the level of modification, we can have conjunction and gapping of NOs, cf. (24), (25). (see also paper of Meya on this topic)

(24) The beautiful house and garden for Susan
(25) The beautiful house and green garden for Susan

(26) The beautiful house of Paul and green garden for Susan

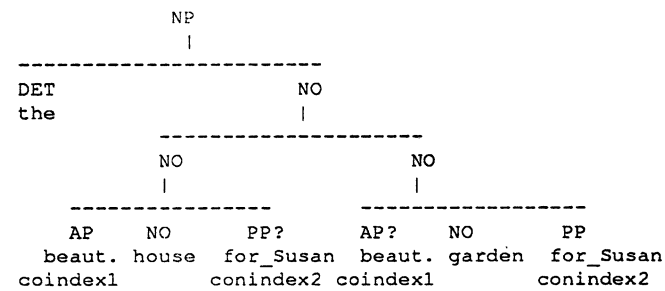
Determination operates in this case over the whole conjunction, the structures of (25) being (25'):

(25')



In this structure, we have scope problems, as indicated in (24'):

(24')



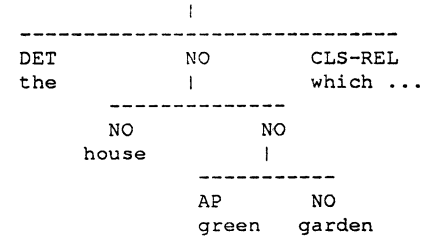
According to the minutes of the Barcelona workshop 89, we decide to take the smallest possible scope, avoiding ambiguities except if it is clear that scope might be different (i.e. in (24') we would NOT copy the AP and PP). Missing nodes are marked with (GAPPED T) in analysis and coindexed with their fully specified counterparts. (There is one exception to this treatment: If genitives are attached, they have scope over both conjuncts, in order to cover the similarity between (27) and (28):

(27) John's beautiful house and green garden
(28) The beautiful house and green garden of John

An effect of this treatment is that we do not have a flat NP structure here but a complex NO structure, and that some of the complements (and modifiers) are not represented at top level but on one level below. But on the other hand, there still can be modifiers which are attached to the top NP level, e.g. CLS-RELs; cf. the agreement behaviour:

(29) The house and green garden which WERE recently bought by him

(29')



In (29) it can be seen that some modifiers may have scope over both conjuncts of an NP. When this is the case is unclear; the tendency should be to represent modifiers on NP level if they cannot be clearly assigned to one of the conjuncts.

3.2.5 Appositions

In this chapter, we have to treat cases like (30) or (31); they are called appositions in the literature:

(30) EN: The man, proud of himself, went into the garden

(31) EN: The man, this idiot, went home
(32) DE: Der Mann, ausgestattet mit einem neuen Gewehr, traf ihn.
(33) DE: Der Mann traf ihn, ausgestattet mit einem neuen Gewehr
(34) EN: Proud of himself, the man came in

These cases seem to be elliptic copula clauses and to have the same positional behaviour as other CLS-SUBs (i.e. come after AVP, PPs etc. in the order of precedence. As they do not have verbs, however, it is difficult to analyse them as clauses (In German, we use a category APPOS which causes problems, however). In MIR, we should state them to be CLS-SUBs with a "KCLS elliptic" marker (or should we attach the GAPPED PRED be here?)

(Alternatives are to treat them as special kinds of PARs, or to treat them as defective CLS-RELs, as EN analysis does right now).

Appositions can be moved in a clause, cf. (33), (34). This causes problems if they should be related to their respective heads. While (33) is elliptic, (34) clearly refers to the $SUBJ NP (cf. Baker 89). The procedure searching for the head is similar to the ATTACH-CLS-REL procedure and should be moved to the INTEGR part of the rules.

3.2.6 Relative Clauses

There are two kinds of relative clauses: restrictive ones should be treated as complements and be marked with (FUNC COMPL) (as they fall into the scope of a potential modifier, cf. (35)), nonrestrictive ones are attached later on and marked as (FUNC MOD). How the two kinds of relatives can be distignuished is language specific: Punctuation (cf. (35)), specific particles (36), etc. In cases where we cannot distinguish the two cases (like in (37)) we should prefer the restrictive reading according to Grice's conversation maxime "be relevant".

(35) EN: the old thief who came late died
the old thief, who came late, died
(36) DE: der Dieb, der uebrigens zu spaet kam, starb
derjenige Dieb, der zu spaet kam, starb
(37) DE: der Dieb, der zu spaet kam, starb

CLS-REL attachment has to be done by special tree searching procedures in the INTEGR part of the grammar (In the G-E system, there is a procedure ATTACHRELCLS which does attachment in the analysis already; as it uses a method different from the one used in anaphor resolution, we sometimes recognise an antecedent different from the actual attachment. To avoid that we need a uniform approach).

3.2.7 APs

There is just one question to be discussed here, namely participila clauses and similar constructions:

(38) DE: das von ihm gekaufte Buch
(39) EN: the book bought by him
the book which has been bought by him

Some analyses deliver a complex AP (marked with PF PAPL), others deliver sentences like CLS-RELs or elliptic CLS-SUBs.

There seems to be no reason to find a common representation of these structures in MIR; as some languages translate APs into APs (german-to-danish), others translate APs into CLS-RELs, others CLS-RELs into CLS-RELs), this should be decided by each transfer module (the worst case is that there are generation rules wich do never fire).

4. Determiners, quantifiers, and numerals

This section is mainly covered by the paper of Meya. The following is just a summary of her paper.

4.1 Quantifiers

In MIR, there will be no SYNTACTIC category QUANT. (Each language is free to use this category, but it should be renamed at the end; e.g. Dutch analysis should have ((CAT NP) (QUANT T)) instead of ((CAT QUANT)(NP T)).

Quantifiers have different syntactic realisations: Adjectives, determiners, XMODs, nouns ("dozen"), etc. These lexical units are marked with three features:

- QUANT T says that this is a quantifier
- QUANT_TYPE says what type of quantifier we have; values are
  NUMeral COLLective DISTRibutive PARTitive
  PROPortional EXCLUSion AMOUNT (see Meya)
- DEFNESS says whether this quantifier includes definiteness;
  values are DEF ("all") INDEF ("some") and
  UNM (unmarked wrt info on defness)

These features are lexical; they are percolated up the tree onto the NP node.

In addition, we have a feature FOCUS with values DEF INDEF and UNM which give the redundant information on the NP for cases like

(1) All the boxes
(2) a few houses

where DEFNESS of the quantifier "overrwrites" the defness information of the determiner. What is indicated here is that these constituents are in focus; and this is marked by the "FOCUS" feature.

Partitives, negatives, and amounts are also treated in the paper of Meya (see there).

4.2 Numerals

Numerals are treated as quantifiers (QUANT_TYPE NUM). They can be modified, cf.

(3) EN up_to 4 houses
(4) DE die ca. 30 Sitze

Therefore we need something like NUMP as a category for NUMs. This holds for cardinals.

4.3 Determiners

Determiners are NOT deleted: Generation parts can do that if they want. However, determination is featurised for languages which do not have an explicit determiner (like Danish sometimes).

Determiners have their KD ("kind of determiner") features as already stated in MIR, with one change: KD QUANT is problematic as QUANT DETs can be definite as well as indefinite; therefore we propose to extract QUANT from this list and use QUANT T and other features (see above).

Moreover, DETs have DEFNESS with DEF INDEF UNM. The problem here is that we have to distinguish between morphological and semantic definiteness, as also generics have definite determiner. DEFNESS therefore is a property with regard to the whole clause (see Meya on generics), not just the local NP. However, presence of DEFNESS is one of the pragmatic clues to determine definiteness.

DETs will have FUNC DET. This also holds for Genitives like (5) which would be marked as (5'):

(5) John's book

(5') NP NO
FUNC DET FUNC HEAD
ROL $SUBJ
DEFNESS DEF

NP
DEFNESS DEF

5. NP modifiers, XMODS

XMODs have the whole NP as their scope. They can have different functions; they can be quantifiers, negations, etc. See paper of Meya on this topic.

6. Special Problems

This section deals with some special types of NPs which have to be recognised by METAL; they should be marked as such in MIR. For this purpose, we need a feature, called SPNP (for "special NP"), its values being AMOUNT, DATE, and others.

6.1 Amount or Measure NPs

(This is an alternative to the proposal of Meya; but we tend to adapt her treatment of measure and amount NPs).
Measure NPs describe amounts of something, cf. (1)-(3). They consist of an amount NP and a dimension.

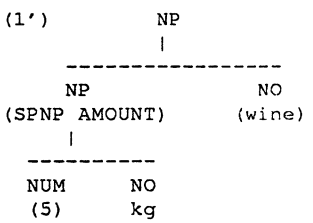
(1) EN: five kg of wheat
(2) DE: 3 Kisten Wein
(3) FR: trois bouteilles de Champagne
(4) DE: er wiegt 4 Tonnen

The amount NP consists of a measure unit ("kg", "bouteille"), which is quantified by a number. This NP can have predicative use (cf. (4)) or attributive use (cf. (1)-(3)). In this case it is attached to a noun which specifies the dimension of the amount; it is mostly a mass noun (in NO complement position, with or without a partitive particle). In NPs, we talk about attributive uses. Amount NPs are marked with the feature (SPNP AMOUNT). (Proposal for terminology: call the amount NP (i.e. the number plus unit) "amount NP"; and call an amount NP plus its dimension a "measure NP")

The syntactic head of these construction is the unit as the agreement shows, cf. (4), (5): The head forms a complete NP, cf. (6); it can be modified, quantified etc.; the second NP is resticted, however: In German, it can be modified (cf. (7)), but not quantified or determined (cf. (8),(9)); in English, it cannot be quantified either (cf. (10)); and if it is determined (cf. (11)), it is a kind of partitive NP (see below 5.2).

(4) EN: Five kg of wheat were stolen
(5) DE: 3 Kisten Wein wurden getrunken
(6) EN: the 5 kg of wine
(7) DE: 3 Kisten guter Wein
(8) DE: * 3 Kisten einiger/viel guter Wein
(9) DE: * 3 Kisten der gute Wein
(10) EN: * 3 bottles of much wine
(11) EN: 3 bottles of the good wine

This shows that the syntactic properties of amount NPs are very similar to other NPs in DET position, as they block additional DET or QUANT nodes within this NP. Therefore, (1) could be analysed as (1') (cf. Bunt 85 for a semantic description of this analysis):



This structure is similar to the genitive NP attachment phenomena. All other constituents could be attached as usual:

(13) EN: 5 kg very good wheat from Alaska

On the other hand, a structure like (1') does not take into account that syntactically, the head of the NP is the amount/unit, as the agreement cases clearly show. Therefore, syntactically, structure (1'') should be preferred:

```
(1'')          NP
                |
       ------------------------
       NUM      NO          NO
        5    (FUNC HEAD)  (FUNC APP)
               kg          wheat
```

This structure would treat the mass noun as a kind of name or identifier (just like (14)):

(14) DE: 3 Systeme METAL

This structure, however, is not able to express the narrower binding of the NUM to its amount (as is shown in the predicative use of amount NPs where they really form proper NPs), and it also does not explain why in (13), we would attach the "Alaska"-PP to "wheat" rather than to "kg".

The source of the problem is, of course, that there is a difference between the semantic and the syntactic head of this construction. We propose to decide that structure (1') should be taken, as it is semantically clearer and captures also some syntactic regularities which (1'') does not. The exception would be to use the NU feature of the amount NP at the top NP (to express the syntactic headedness and agreement behaviour).

There is one problems left: How to recognise amount NPs? In METAL, we could use nouns marked as TYN MEA to be heads of amount NPs. But as MEA is a function of a noun rather than a property, and as many nouns can be used in this function (in fact, by putting them into this "amount context"), this is only part of a solution. However, the problem comes from the inconsistent use of semantic features rather than a conceptual problem: We should have readings of nouns, and a reading can consist of several features (see 6 below): It can specify the function MEA together with its dimension (e.g. TMP, LOC etc.), i.e. what kind of measure it is. Then we can identify an amount NP by a number (or similar constructs like "dozen") and a MEA noun. (This implies coding of units like "kg", "m" etc. as NSTs, ABBs, and (TYN MEA)s).

6.2 Partitive NPs

They are described in the paper of Meya.

The basic idea is that the NUM (or quantifier) acts as head of the constituent, and it has a COMPL PP attached to it:

```
(15)   five of the books
       NUM     PP
         |
        PRN
        ---------
           NP
```

(This could be problematic in cases like (16) where the NUM is modified:

(16) GE: die besten fuenf von den Buechern)

The partitives are marked as QUANT_TYPE PART.

(How does FR treat partitive structures like "du lait" etc.?)

(Eurotra treats partitives as gapped constructions where the head of the PP is considered to have been gapped in the NP position: "five [books] of the books". Is this an alternative?)

6.3 Terms, Idioms, Multiwords

As explained above in the compound section, multiwords are semantic units with a syntactic structure. This structure must be analysed and

generated.

If a multiword is lexicalised, it might be a lexicon entry in a mono lexicon (problem: internal inflection). If it is not, its transfer can be computed. There are two possibilities: either transfer is completely regular, then there are standard means of transfer (e.g. a DE compound specifier is transformed into a SP "de"-PP). Or some lexical changes have to be made. This has to be stated in the transfer lexicon.

6.3.1 Analysis of multiwords

The basic idea is that if a multiword is not lexicalised, its analysis is done just as any other syntactic analysis. I.e. the system does not "know" that it analyses a multiword. We assign just ordinary syntactic structures: (17) will be a NO with a PP, (18) will b an AP with an NO, etc.

(17) FR machine a laver -> GE Waschmaschine
(18) DE logische Verbindung -> EN session

It is only in transfer that a multiword is recognised as such, and a special translation must be triggered.

In the proposal of Vazquez 88, potential multiword parts are marked by nominal framing, roles like $ADJ and $POBJ are used. This is not quite correct, however, as noun complements are something different (e.g. "logisch" would never be a "complement" of "Verbindung"); on the other hand, there are only the complements/modifiers near the left or right of the head involved. Marking them does not need framing software, this can be done by pure syntactic means (marking the respective APs or PPs next to the heads).

In addition, as we could have transfer only at NST level, all relevant information had to be copied down to the terminal node of the head in order to be accessible by the transfer procedure. This could be replaced by a more general principle that the controller (i.e. the head) of the NP is transferred first, and it controls the transfer of the other constituents. This procedure is similar to the transfer of PREDs on CLS level.

6.3.2 Transfer problems

In transfer, we could have tests and actions like the following:

- test for the existence of a specific modifier, e.g. AP with CAN "logisch" and head with "TYN CNC"
- test for a PP with CAN "a" and head "laver"

etc.; the relevant operations have to be specified in more detail.

Then, we could have actions which create a compound specifier (CAN "waschen"), which LEX-DELETE the "logisch", which insert a preposition, etc.

Just as in the PRED transfer, these actions are triggered if the tests are successfull, and the transferred CAN and CAT are passed down the NP head line.

6.3.3 monolingual idioms

If a multiword is lexicalised, it has to be treated as one unit already in the analysis. This is a problem of idioms. We will assume that idioms are just strings of more than 1 word (i.e. have a space in between). Problems of internal inflection or of inserted other constituents are not considered at present.

In these cases, we have the problem that there may be competing readings: idiom vs. compositional analysis, both of which can be valid, cf.

(19) up to 5 books (up_to is XMOD for NUMs)
(20) he got up to give a talk

In DE, we had an idiom filter which systematically blocked the

compositional reading by removing it from the chart; i.e. (20) would have been a phrasal. This is certainly not the best way. With the possibility of lexical PRFs (and the default idiom PRF 3), both readings can be accepted and be tuned to a certain extent. Even then, however, idiom readings (and transfers) are not picked although they were coded (which is difficult to explain to customers).

Transfer is not a problem in these cases: SL CAN is quite simple (it is the idiom); and either we can assign a TL CAN immediately (cf. (21)), or we can construct a complex transfer expression, using XFMS for ADDing lexical material, cf. (22). (This could be named procedures which either construct strings or PPs or whatsoever; ADD-STRING is a step into that direction). This way, we could achieve the inflection of the head of the new constituent.

```
(21) EN up_to  XMOD   ->  DE  bis_zu  XMOD
(22) DE Maschine  NST ->  FR  machine NST
     Test: (FUNC SPEC), (CAN waschen) ...
     XFMS: ADD (CAT PP) (ROL $POBJ) (ALO a_laver) ...
```

These operations should be done on NP level, of course.

6.4 Dates

Dates are a special type of NP and have to be treated separately (as they are units of time as a whole, like multiwords. But they have to be translated differently in the different languages.

Dates are recognised by the respective analysis modules. They are marked with (SPNP DATE) and receive a TYN value (before framing!).

They should have an internal MIR structure. The date itself should be dominated by a NO node (as dates can be determined, quantified, etc.) and consist of an ordinal AP, a month as head and a year as apposition. This might be enough to determine the transfer. (We could go further and consider e.g. "5" to be ((CAN may) (ALO 5) (ABB T)) but this might not be necessary).

6.5 Gerund structures in English

Are they NPs? Or should they be treated as N0, i.e. as clausal fillers of these positions: Is (23) N1 or N0? What about (24)? cf. (25) which is a nominalisation (NVRB T), (26):

(23) Reading books is beautiful
(24) His reading books is beatutiful
(25) The reading of books is beatutiful
(26) The reading books is beautiful

What MIR structure should we produce?

6.6 Pronouns

The function like NPs, i.e. they cannot be determined. They can have modifiers, however, cf. (27):

```
(27) EN Somebody from Hamburg went in
(27')       PRN
            NP      PP
        FUNC HEAD FUNC MOD
        ------------------
                NP
```

In this case, we would have to repeat some of the MOD-attaching rules at NP level, and have a structure like (27') for them (where the head of the NP is an NP again).

7. NP Semantics

This chapter deals with a proposal for the treatment of the semantic features in NPs.

In METAL, we have a set of several semantic features for nouns, adjectives and other constituents (prepositions etc.). Features are used to control transfer, but things become more complicated now.

As Behrend/Wienhard 89 show, features are used inconsistently in METAL:

a. Disambiguations in transfer do not have their TL counterparts; no notion of a "reading" is used in METAL.

b. Sets of features sometimes describe one reading, sometimes a real ambiguity (i.e. several readings).

c. Implications are not made explicit in METAL: A noun with TYN HUM will not passs a transfer test for TYN ANI.

d. No calculation of semantic types is done: We would attach a colour ADJ to an abstract noun without any problem.

e. There are interrelations between framing and NP semantics: Even if an ambiguous NP could be disambiguated (because framing decided on its type), this information is not used in the NP transfer, and a wrong transfer still could be picked.

Therefore we propose to proceed as follows:

1. We will introduce readings into the METAL semantics. Transfer will be done between readings. Readings have to have their TL counterparts.

2. Readings are described by sets of features. More than one feature can be used to describe a meaning, e.g. TYN (CNC MI) describes a concrete entity which also is machine intervention, TYN (MEA TMP) describes a temporal measure (like "month"), etc.

3. The introduction of readings should not lead to more ambiguous analyses (e.g. 3 parses for a semantically ambiguouos word). Therefore we propose to keep the readings in one mono entry if they are not syntactically different. This means that the structure of the TYN feature becomes more complex: It will be a list of lists now. E.g. "Bank" (ABS-> bank CNC-> bench) will have (TYN (ABS) (CNC)) instead of (TYN ABS CNC), indicating that there are two competing readings.

4. Once we can distinguish between readings, we can manipulate them. We need computation of semantics esp. when attaching APs and PPs, and when framing is done. Some readings could be excluded. This could be done by named procedures (which seem to be universally useable). They will work basically like constraints in the old Katz/Fodor sense but without amalgamation: They just exclude implausible combinations.

5. In transfer, tests are performed between readings. Here we should have different transfers according to different readings. If more than one reading survived the analysis filters, all possible transfers for all of these readings should be given (At present, the system seems to randomly pick the first one) as long as there is no other criterion (which e.g. could be a requirement of the TL verb frame). On the other hand, something like a default should be kept.

This has to be clarified in more detail, but it seems that this is a sounder basis for the semantic features system. At present, there is much semantic information in the system (TYN, TYA, P1/P0 etc.) but it is badly organised and not used efficiently.

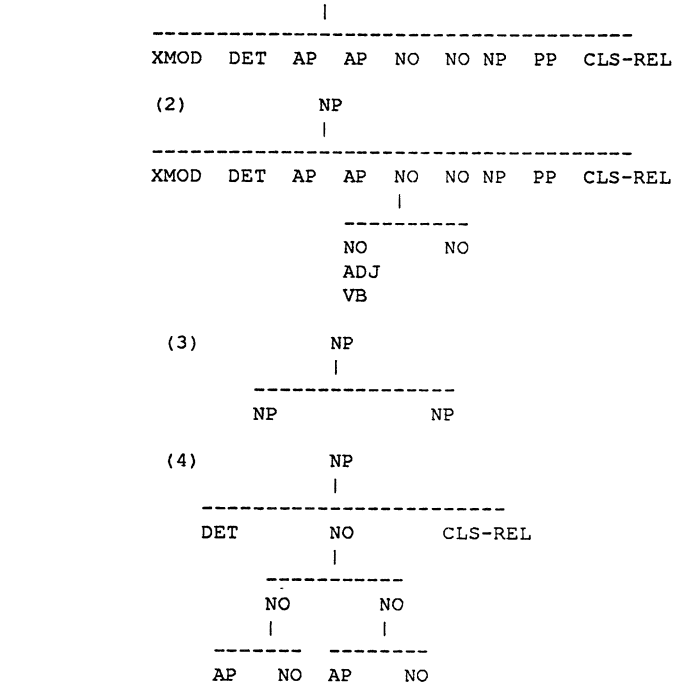
8. Structures, features, and comparison

Different analysis parts should be converted into a common MIR structure. This structure looks as follows:

8.1 Basic NP structure

The structure of a standard NP is flat. Precedence is oriented on the surface precedence; no canonical order is given. Standard structure is (1); compounds have complex NOs (cf. (2)); conjunctions are done on NP and NO level (cf. (3) (4)).

(1) NP



I.e. the head NO is complex in cases of conjunctions (marked with CNJ T) or compounds (marked with CMPD T).

8.2 NP features

NU	SG PL	number of NP / NO
GD	M F N	gender of NP / NO
CA	N G D A	case of NP / NO
CMPD	T	if a NO is a compound (NO)
HEAD	(CAN)	set at FUNC SPEC within compounds
CONST	T	if a NO is a CONSTANT
NADJ	T	if a NO is a nominalised adjective
NADV	T	if a NO is a nominalised adverb
NVRB	T	if a NO is a nominalised verb
ABB	T	if a NO is an abbreviation
DEVERBAL	T	if a noun is deverbal ?
DG	POS COM SUP	degree of NO / NP
COMCON	T	NP has a comparative construction inside
PO	BEF AFT	if a constituent is before or after the head
ARGS	$SUBJ $DOBJ $POBJ $SOBJ $TMP $LOC $MEA	arguments of a noun
ROL	$SUBJ $DOBJ $POBJ $SOBJ $TMP $LOC $MEA	realised roles in a NP
FUNC	HEAD SPEC APP COMPL MOD DET	function of a constituent within an NP
TYN	ABS CNC LOC ...	semantic type of the NO / NP
KN	MASS COUNT	kind-of-noun
...		(all other NST features except CL)
KD	DEF IND POS INT DEM REL	kind of determiner
QUANT	T	constituent is a quantifier
QUANT_TYPE	NUM COLL DIST PART PROP EXCL AMOUNT	type of quantifier
DEFNESS	DEF INDEF UNM	(semantic) definiteness
NEG	T	negation
GENERIC	T	generic NP

FOCUS	DEF INDEF UNM	NP in focus with defness = (value)
CNJ	T	if a NO / NP is conjoined
GAPPED	T	if a NO / NP node has been gapped
CNJCAN	(CAN)	CAN of the conjunction
SPNP	DATE PRONNP GND	special NP (date, pronoun head, gerund)

8.3 Transfer

Transfer will be done at NP level, triggered by a special call similar to FRX. It will transfer the head of the NP, using etsts and XFMs just like PREDs. The result of the transfer will be passed down via the TRF feature, just like PREDs again.

Transfer will test for complements, modifications, multiwords, compound elements etc., and it will have actions like adding, deleting and changing lexical material.

9. References

- Vazquez, B.: Nominal Framing and Multiwords (3rd METAL Workshop 1988)

- Neunzig, W., Grauwinkel, G.: Zum Problem der Nominalkompositalexikalisierung in METAL (3rd METAL Workshop 1988)

- Gebruers, R.: METAL Bible, Coding Principles. 1989

- Khan, R., Oppenheim, J., 89 (Morphology) (Proc. Barcelona Workshop 89)

- Wiesner, M., Jezierny, L.: (Abbreviations) (Proc. Barcelona Workshop 89)

- Adriaens, G.: The self-extending lexicon, Off-line and On-line Defaulting (Proc. COLING 90)

- Liu, J.: NP structures in MIR. 1989

- Baker, J.: English Grammar. 1989

- Bunt, H.: Mass Nouns and model theoretic Semantics. 1985

- Thurmair, G.: (CONST, ID) (Proc. Barcelona Workshop 89)

- Behrend, L., Wienhard, H.: 4. Zwischenbericht zu "Disambiguierung lexikalischer Einheiten". 1989